Network models – random graphs

Properties common to many large-scale networks, independently of their origin and function:

1. The degree and betweenness distribution are decreasing functions, usually power-laws.
2. The distances scale logarithmically with the network size

$$l \approx \frac{\log N}{\log\langle k\rangle}$$

3. The clustering coefficient does not seem to depend on the network size

$$C \propto \langle k\rangle$$

As though all these networks were part of the same family/class.

Random networks

The average distance and clustering coefficient only depend on the number of nodes and edges in the network.

This suggests that general models based only on the number of nodes and edges in the network could be successful in describing the properties of an “expected” (characteristic) network.

Uniformly random network: distributes the edges uniformly among nodes.

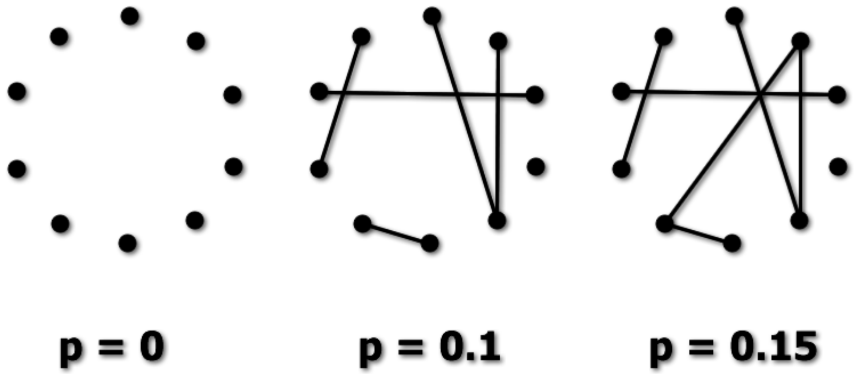
Probabilistic interpretation:

There exists a set (ensemble) of networks with given number of nodes and edges. Select a random member of this set.

What are the expected properties of this network? – studied by random graph theory.

Random graph theory

Erdös-Rényi algorithm - Publ. Math. Debrecen 6, 290 (1959)



- fixed node number N
- connecting pairs of nodes with probability p

Expected number of edges: $$E = p\frac{N(N-1)}{2}$$

Random graph theory studies the expected properties of graphs with $N \to \infty$

The properties of random graphs depend on *p*

Properties studied:

is the graph connected?
does the graph contain a giant connected component?
what is the diameter of the graph?
does the graph contain cliques (complete subgraphs)?

Probabilistic formulation: what is the probability that a graph with N nodes and connection probability p is connected?
Increase p from 0 to 1. Some of these properties appear suddenly, at a threshold p_c

$$P_{N,p}(Q) = \begin{cases} 0 & \text{if } p < p_c \\ 1 & \text{if } p > p_c \end{cases}$$

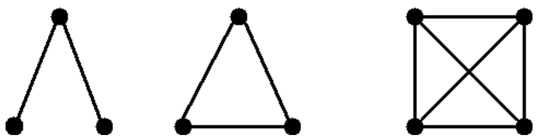
Note that p_c depends on N.

Critical thresholds for the emergence of certain subgraphs

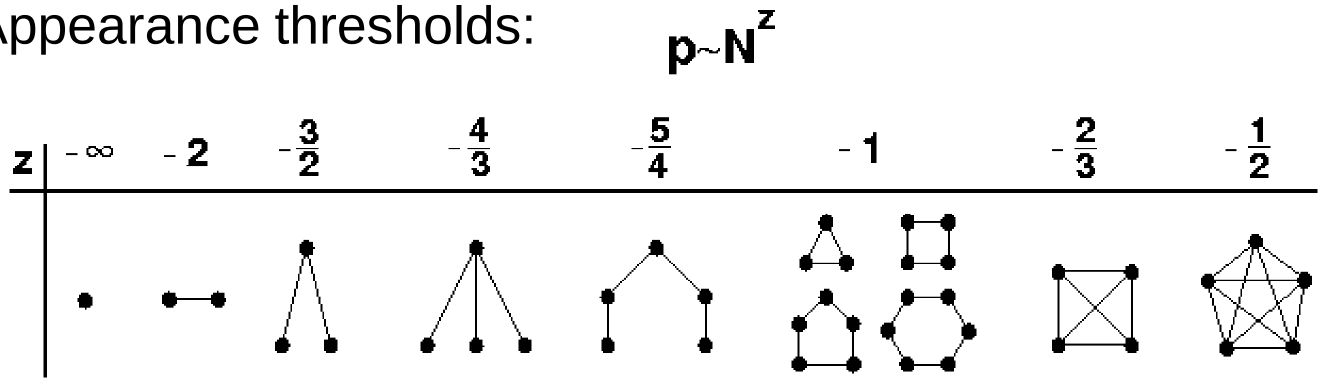
Assume that the connection probability is a power-law of N, $p = cN^{z}$

Assume that z increases from $-\infty$ to 0

Look for trees, cycles and cliques in the graph.



Appearance thresholds:

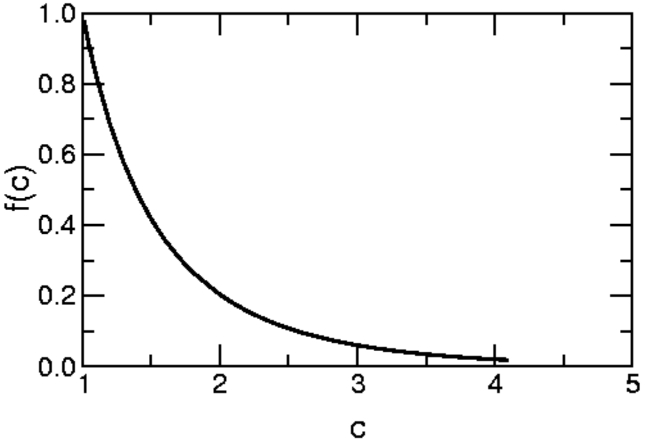


The graph contains cycles of any length if $z \geq -1$

Emergence thresholds for clusters in a random graph

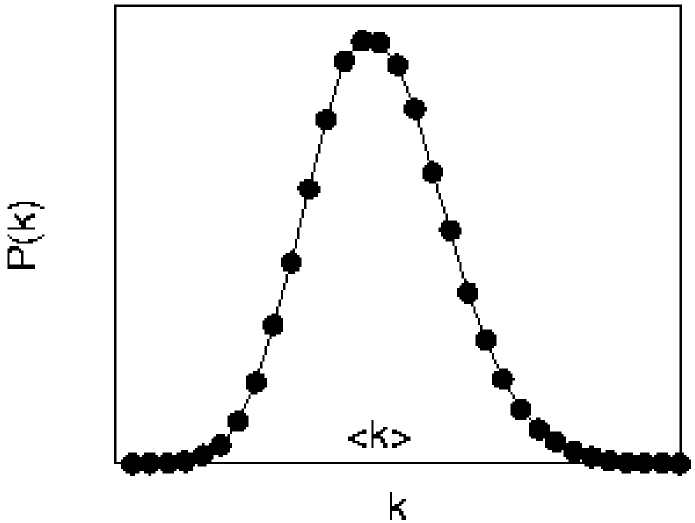
- For $p < N^{-1}$ the graph contains only isolated trees.
- If $p = cN^{-1}$ with $c < 1$ the graph has isolated trees and cycles.
- At $p = cN^{-1}$ with $c = 1$ a giant connected component appears.
- The size of the giant connected component approaches N rapidly as c increases.

$$S = (f(1) - f(c))N$$



- The graph is connected if $p > \ln N / N$

Node degrees in random graphs



- average degree: $\langle k \rangle = \dfrac{2E}{N} \cong pN$
- degree distribution:

$$P(k) \cong C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

ways to select *k* nodes from *N-1*

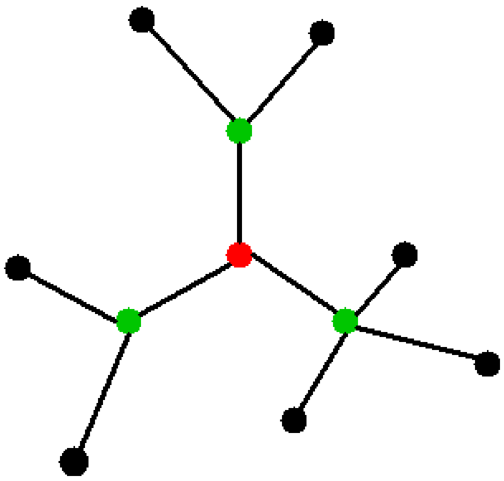
probability of having *k* edges

probability of missing *N-1-k* edges

Most of the nodes have approximately the same degree.
The probability of very highly connected nodes is exponentially small.

Distances in random graphs

Random graphs tend to have a tree-like topology with almost constant node degrees.



- nr. of first neighbors: $N_1 \cong \langle k \rangle$
- nr. of second neighbors: $N_2 \cong \langle k \rangle^2$
- estimate maximum distance:

$$1 + \sum_{l=1}^{l_{max}} \langle k \rangle^i = N \quad \Rightarrow \quad l_{max} = \frac{\log N}{\log \langle k \rangle}$$

This scaling was proven by Chung and Lu, Adv. Appl. Math 26, 257 (2001).

There is no local order in random graphs

Clustering coefficient: $$C_i \equiv \frac{n_i}{k_i(k_i-1)/2}$$

Since edges are independent and have the same probability *p*,

$$n_i \cong p\frac{k_i(k_i-1)}{2} \Rightarrow C \cong p = \frac{<k>}{N}$$

The clustering coefficient of random graphs is small.

Are real networks like random graphs?

As quantitative data about real networks becomes available, we can compare their topology with that of random graphs.

Starting measures: N, <k> for the real network.

Determine *l*, C and P(k) for a random graph with the same N and <k>.

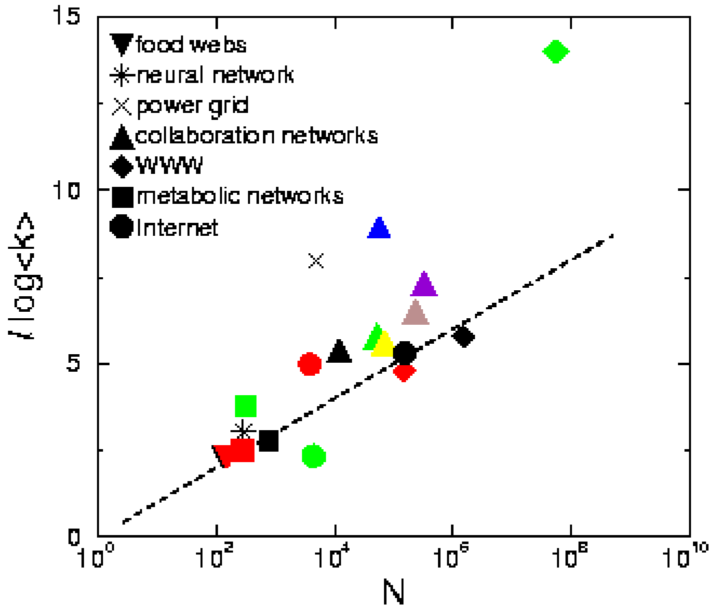
$$l_{rand} \approx \frac{\log N}{\log\langle k\rangle} \qquad C_{rand} = p = \frac{\langle k\rangle}{N}$$

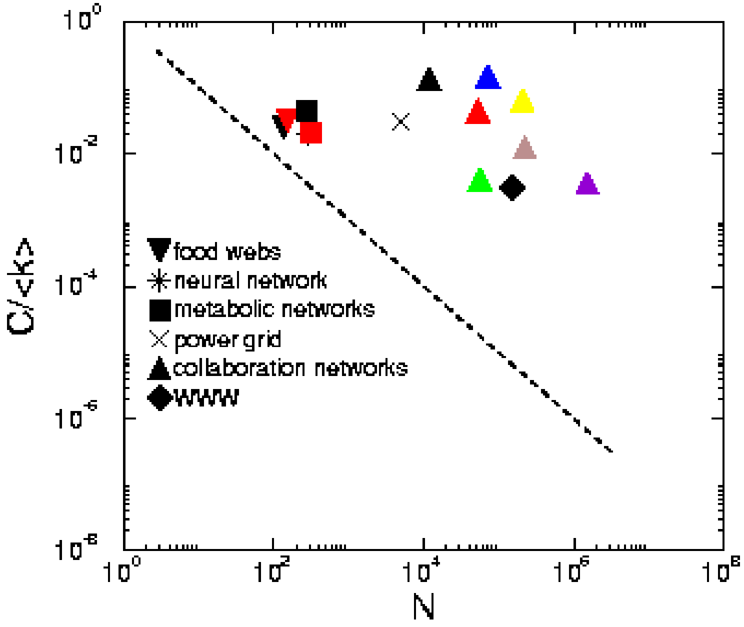
$$P_{rand}(k) \cong C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

Measure *l*, C and P(k) for the real network. Compare.

Path length and order in real networks

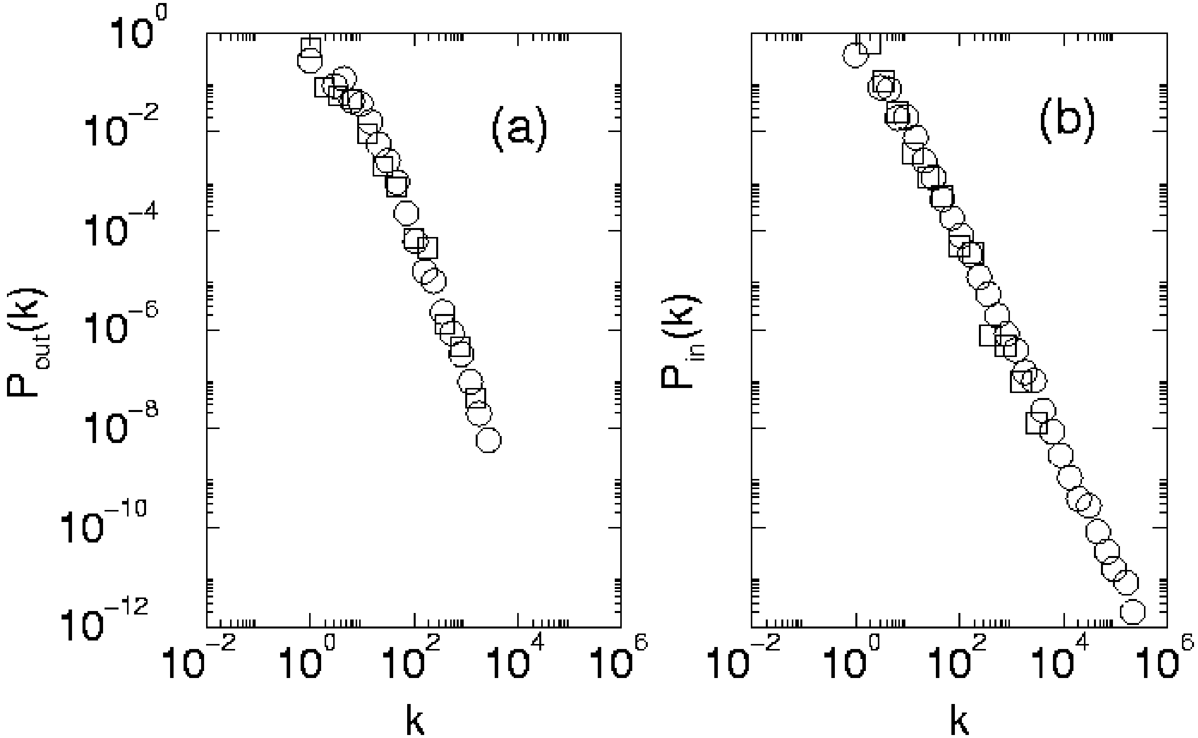
$$l_{rand} = \frac{\log N}{\log\langle k\rangle} \qquad C_{rand} = \frac{\langle k\rangle}{N}$$





Real networks have short distances like random graphs but they are more transitive.

The degree distribution of the WWW is a power-law



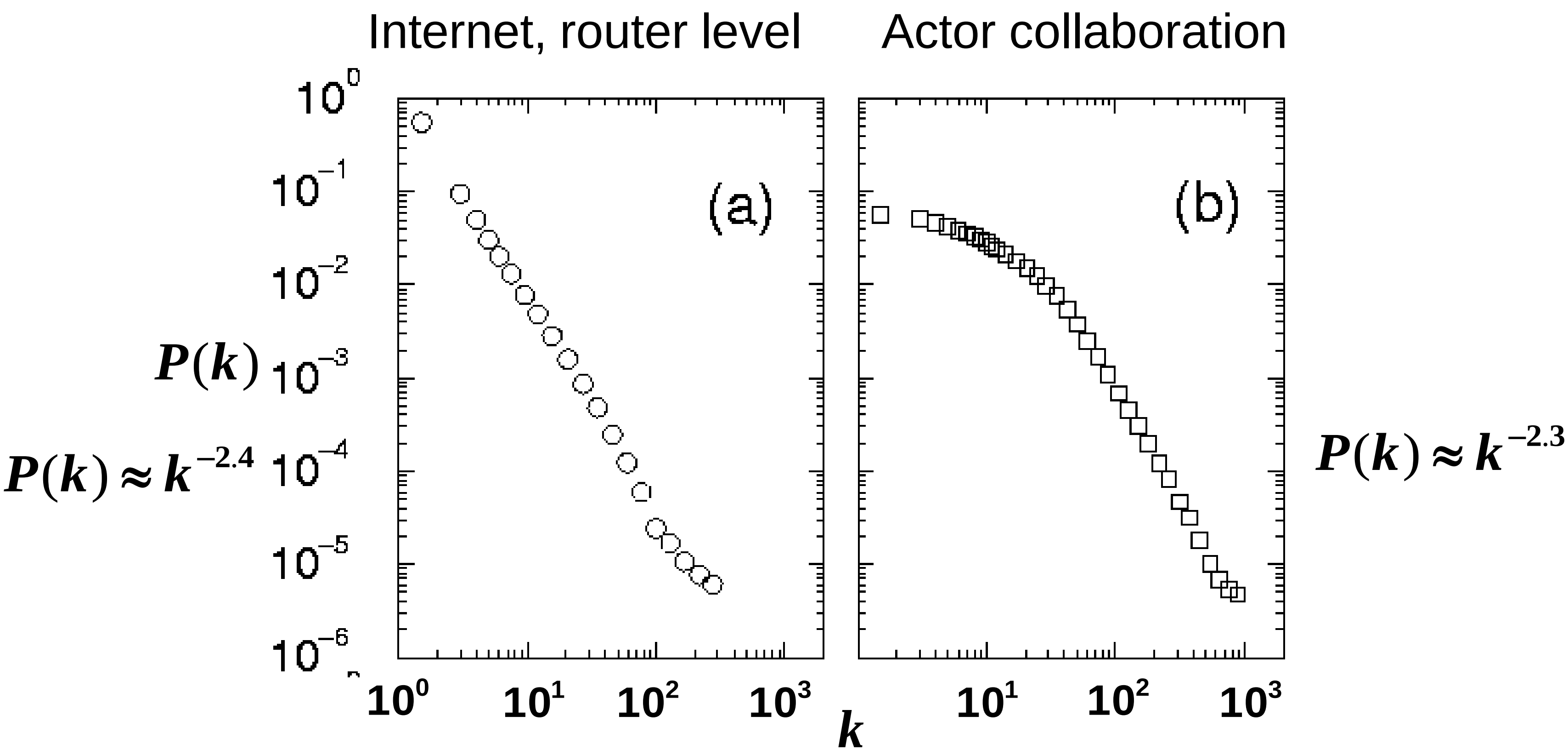
$$P_{out}(k) \approx k^{-2.45}$$

$$P_{in}(k) \approx k^{-2.1}$$

R. Albert, H. Jeong, A.-L. Barabási, Nature 401, 130 (1999)

A. Broder *et al*., Comput. Netw. 33, 309 (1999)

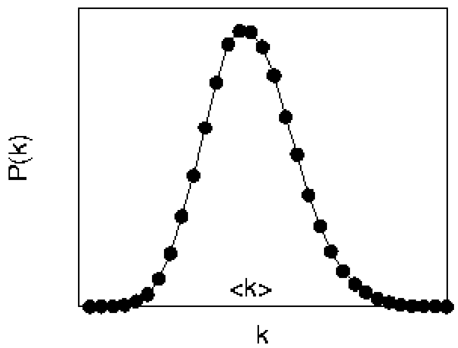
Power-law degree distributions were found in diverse networks

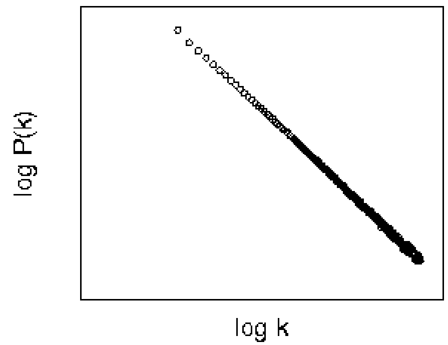


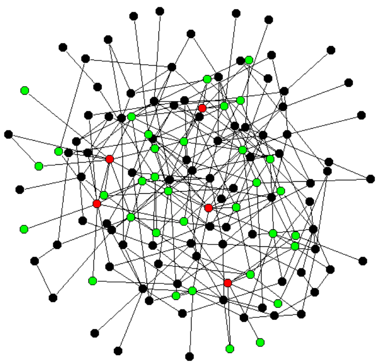
R. Govindan, H. Tangmunarunkit, IEEE Infocom (2000)

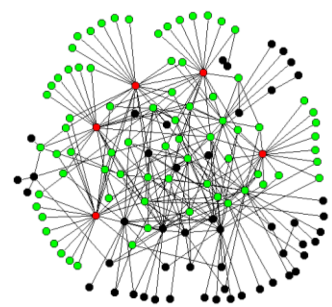
A.-L. Barabási, R. Albert, Science 286, 509 (1999)

The power-law degree distribution indicates a heterogeneous topology









The average degree gives
 the characteristic scale (value)
of the degree.

Large variability,
 the average degree not informative,
 no characteristic scale for the degree

Scale-free

Random graphs with a power-law degree distribution

Fixed N, $P(k) = Ak^{-\gamma}$, $k < K$

Network assembly - random edges, but enforcing the right $P(k)$

Configuration model:
- choose a degree sequence N(k)=N P(k)
- give the nodes k “stubs” according to N(k)
- connect stubs randomly

M. E. J. Newman, S. H. Strogatz, and D. J. Watts,
Phys. Rev. E 64, 026118 (2001)

Ex. Construct a graph with 10 nodes and degree sequence
N(1)=4, N(2)=3, N(3)=2, N(4)=1.
What is a necessary condition for the graph construction?

Properties of scale-free random graphs

Fixed $N, \quad P(k) = Ak^{-\gamma}, \; k < K$

The graph will have a giant connected component if $\gamma \leq 3.47$

Connected if $\gamma \leq 2$

The average path length scales approximately logarithmically with the number of nodes.

The agreement with real networks' path length is not better than ER random graphs'.

The clustering coefficient depends weakly on N and can even increase if $\gamma < 7/3$

In many applications scale-free random graphs are a better benchmark than ER random graphs.

Are generative network models more successful than random graphs?

Network models

Properties common to many large-scale networks, independently of their origin and function:

1. The degree and betweenness distribution are decreasing functions, usually power-laws. scale - free
2. The distances scale logarithmically with the network size

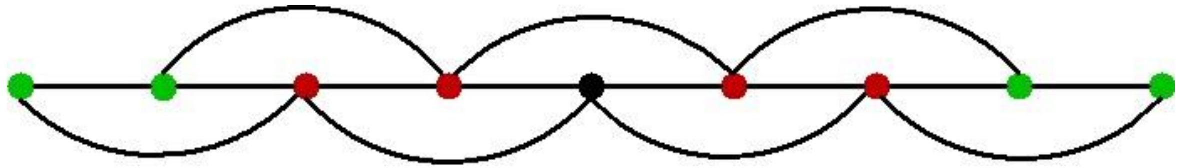
$$l \approx \frac{\log N}{\log\langle k\rangle}$$

small world

3. The clustering coefficient does not seem to depend on the network size, and is larger than the clustering coefficient of comparable random graphs

There are two model families proposed to explain these properties: Small world network models and scale-free network models.

Benchmark: 1D lattice (ring)



$k = 4$ **for each node**

$C = \frac{1}{2}$ **for each node if N > 6**

$$1 + \sum_{l=1}^{l_{max}} 4 \approx N \;\Rightarrow\; l_{max} \approx \frac{N}{4} \qquad < l > = \frac{4\sum_{l=1}^{l_{max}} l}{N} \;\Rightarrow\; < l > \approx \frac{N}{8}$$

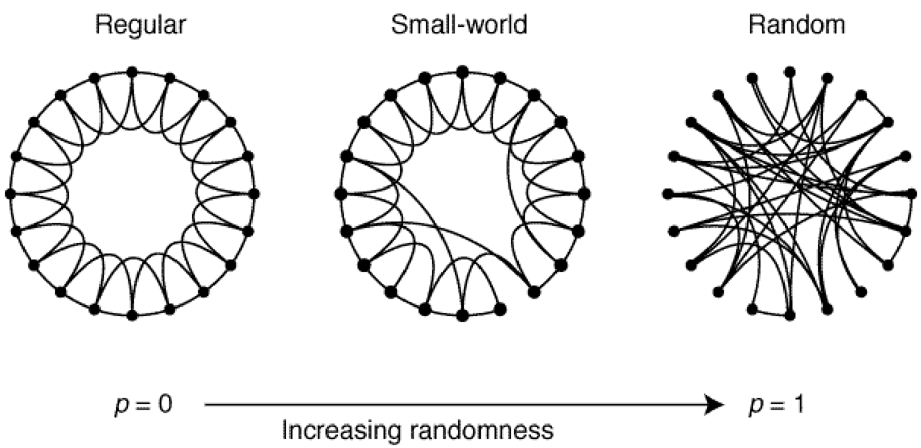
The average path-length varies as $< l > \approx N$

Constant degree, constant clustering coefficient.

Watts-Strogatz model of small-world networks

Real networks resemble both regular lattices and random graphs – perhaps they are in between.

Watts-Strogatz model - D. Watts, S. Strogatz, Nature 393, 440 (1998)

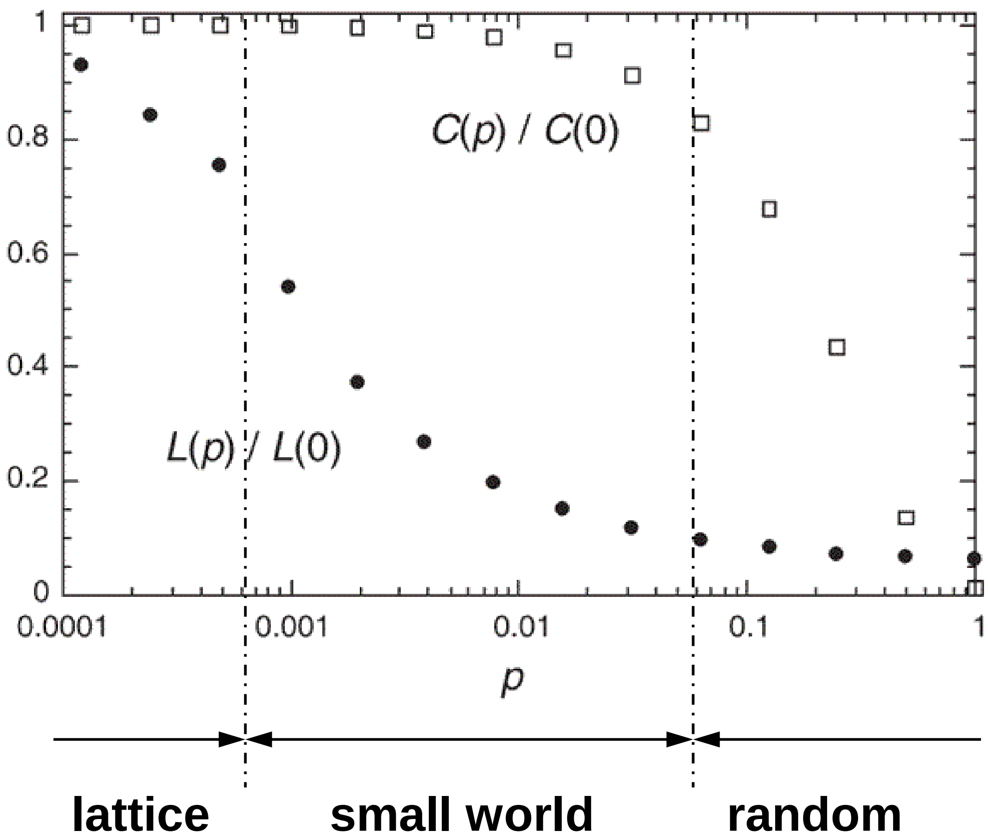


- lattice with K neighbors
- rewire edges with probability p

$$l = \frac{N}{2K}, \; C = \frac{3(K-2)}{4(K-1)} \quad \Rightarrow \quad l \approx \frac{\log N}{\log K}, \; C \approx \frac{K}{N}$$

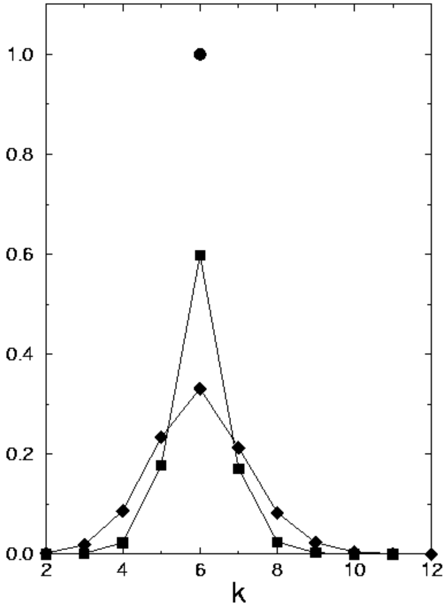
Is there a regime with small l and large C?

Transition from a lattice to a small world



There is a broad interval of p over which $C(p) \cong C(0)$ but $l(p) \cong l(1)$

Degree distribution of a small-world network



Rewiring does not change the average degree, but modifies the degree distribution.

$$\langle k \rangle = K$$

P(k) depends on the rewiring parameter *p,* but is always centered around <*k*>*.*

Degree distribution similar to that of a random graph, with exponentially small probability for very highly connected nodes*.*

We need to uncover the mechanisms responsible for the scale-free *P(k)*

- random graphs
- small-world networks
- scale-free random graphs

Static (number of nodes fixed)

Real networks continuously change

- random graphs
- small-world networks

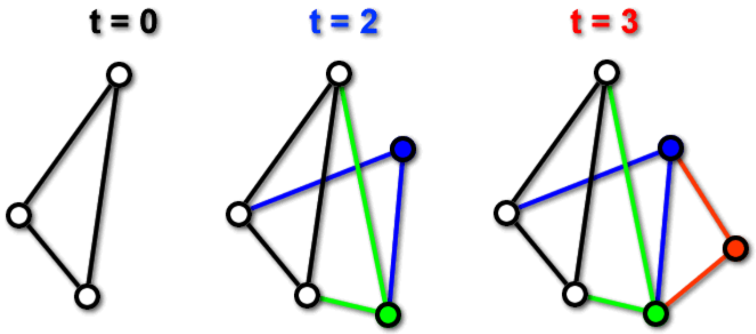
Homogeneous

Scale-free degree distribution - the nodes are not equivalent

We need to model the evolution of networks, not just their topology.

Barabasi-Albert model of scale-free networks

Start with a small seed of m_0 nodes and $m_0(m_0-1)/2$ edges.



- **growth**: a node and m edges added at every step
- **preferential attachment**: $$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$

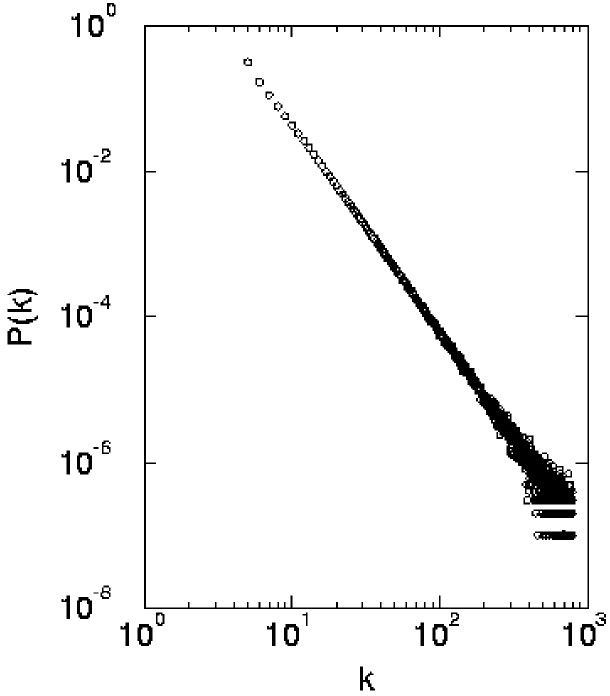
Price, J. Amer. Soc. Inform. Sci. 27, 292 (1976)

Barabási and Albert, Science 286, 509 (2000)

General properties of the network

- nr. of nodes: $N = m_0 + t$
- nr. of edges: $E = \frac{m_0(m_0 - 1)}{2} + mt$
- average degree: $\langle k \rangle = \frac{2E}{N} \rightarrow 2m$
- degree distribution:

$$P(k) \xrightarrow[t \to \infty]{} Ak^{-3}$$



Although the network grows, the degree distribution becomes stationary.

Ex. 1
Start from a seed of two nodes connected by an edge. At each step, add a new node, and connect it by a single edge to a preexisting node.

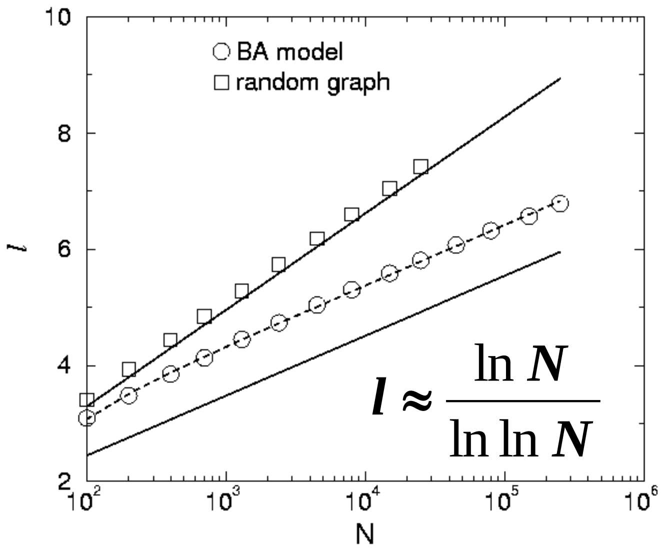
Let the probability of selection be directly proportional with the degree of the “old” node. (Is there an easy way to do this?)

Continue growing the graph until you reach 15 nodes. Describe the graph (average degree, degree distribution,
clustering coefficient, connectivity, maximum distance).

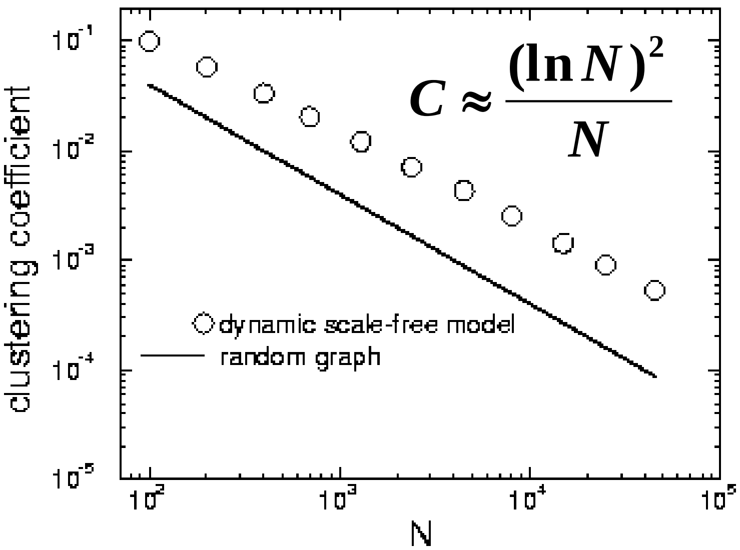
Ex. 2
How will the properties of the graph change if at each step a new node and two new edges are added?

How do the other network measures compare with real networks?

Average path length



Clustering coefficient



Average distances smaller in the BA model than in equivalent random graphs. but not as small as in scale-free random graphs.

Cohen et al, in Handbooks of Graphs and Networks (2003)

Clustering coefficient decreases with network size.

B. Bollobás and O. Riordan, in Handbooks of Graphs and Networks (2003)

Evolving network models

The scale-free model is only a minimal model and does not capture several features of real networks.

Its basic mechanisms can be augmented by the incorporation of

- addition of edges without new nodes
- edge rewiring, removal
- node removal
- constraints or optimization principles

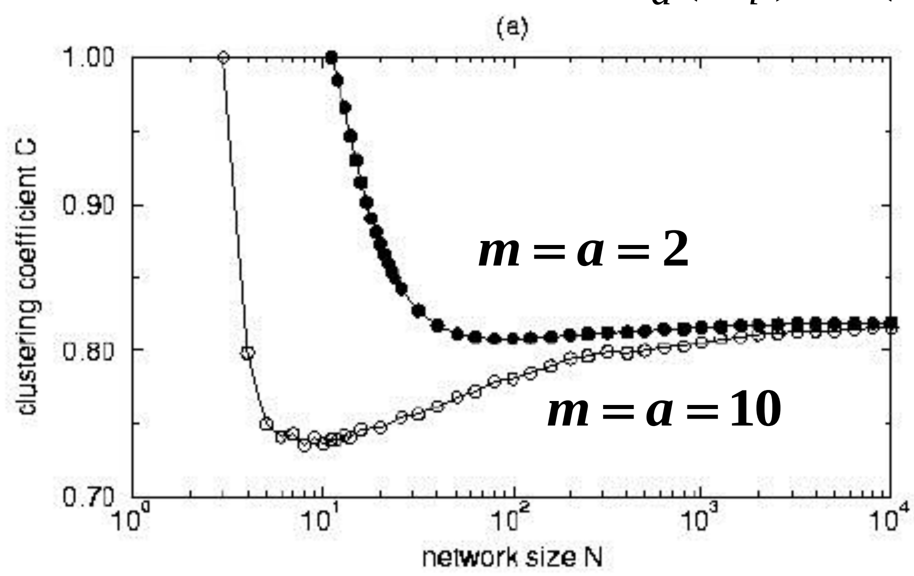
A considerable variety of such evolving network models still lead to heterogeneous topologies, some with close agreement with real networks.

Linear growth, linear pref. attachment	$\gamma=3$	Barabási and Albert, 1999
Nonlinear preferential attachment $\Pi(k_i)\sim k_i^{\alpha}$	no scaling for $\alpha\neq 1$	Krapivsky, Redner, and Leyvraz, 2000
Asymptotically linear pref. attachment $\Pi(k_i)\sim a_\infty k_i$ as $k_i\to\infty$	$\gamma\to 2$ if $a_\infty\to\infty$ $\gamma\to\infty$ if $a_\infty\to 0$	Krapivsky, Redner, and Leyvraz, 2000
Initial attractiveness $\Pi(k_i)\sim A+k_i$	$\gamma=2$ if $A=0$ $\gamma\to\infty$ if $A\to\infty$	Dorogovtsev, Mendes, and Samukhin, 2000a, 2000b
Accelerating growth $\langle k\rangle\sim t^{\theta}$ constant initial attractiveness	$\gamma=1.5$ if $\theta\to 1$ $\gamma\to 2$ if $\theta\to 0$	Dorogovtsev and Mendes, 2001a
Internal edges with probab. p Rewiring of edges with probab. q	$\gamma=2$ if $q=\dfrac{1-p+m}{1+2m}$ $\gamma\to\infty$ if $p,q,m\to 0$	Albert and Barabási, 2000
c internal edges or removal of c edges	$\gamma\to 2$ if $c\to\infty$ $\gamma\to\infty$ if $c\to -1$	Dorogovtsev and Mendes, 2000c
Gradual aging $\Pi(k_i)\sim k_i(t-t_i)^{-\nu}$	$\gamma\to 2$ if $\nu\to-\infty$ $\gamma\to\infty$ if $\nu\to 1$	Dorogovtsev and Mendes, 2000b
Multiplicative node fitness $\Pi_i\sim\eta_i k_i$	$P(k)\sim\dfrac{k^{-1-C}}{\ln(k)}$	Bianconi and Barabási, 2001a
Edge inheritance	$P(k_{in})=\dfrac{d}{k_{in}^{\sqrt{2}}}\ln(ak_{in})$	Dorogovtsev, Mendes, and Samukhin, 2000c
Copying with probab. p	$\gamma=(2-p)/(1-p)$	Kumar *et al.*, 2000a, 2000b
Redirection with probab. r	$\gamma=1+1/r$	Krapivsky and Redner, 2001
Walking with probab. p	$\gamma\simeq 2$ for $p>p_c$	Vázquez, 2000
Attaching to edges	$\gamma=3$	Dorogovtsev, Mendes, and Samukhin, 2001a
p directed internal edges $\Pi(k_i,k_j)\propto(k_i^{in}+\lambda)(k_j^{out}+\mu)$	$\gamma_{in}=2+p\lambda$ $\gamma_{out}=1+(1-p)^{-1}+\mu p/(1-p)$	Krapivsky, Rodgers, and Redner, 2001

A model with high clustering coefficient

- Each node of the network can be either active or inactive.
- There are only *m* active nodes in the network at any instance.

1. Start with *m* active, completely connected nodes
2. At each timestep add a new node (active) that connects to *m* active nodes.
3. Deactivate one active node $P_d(k_i) \propto (a + k_i)^{-1}$



$$\Pi(k) \approx a + k$$

$$P(k) \approx k^{-2-a/m}$$

K. Klemm and V. Eguiluz, Phys. Rev. E 65, 036123 (2002)

Evolving protein interaction networks

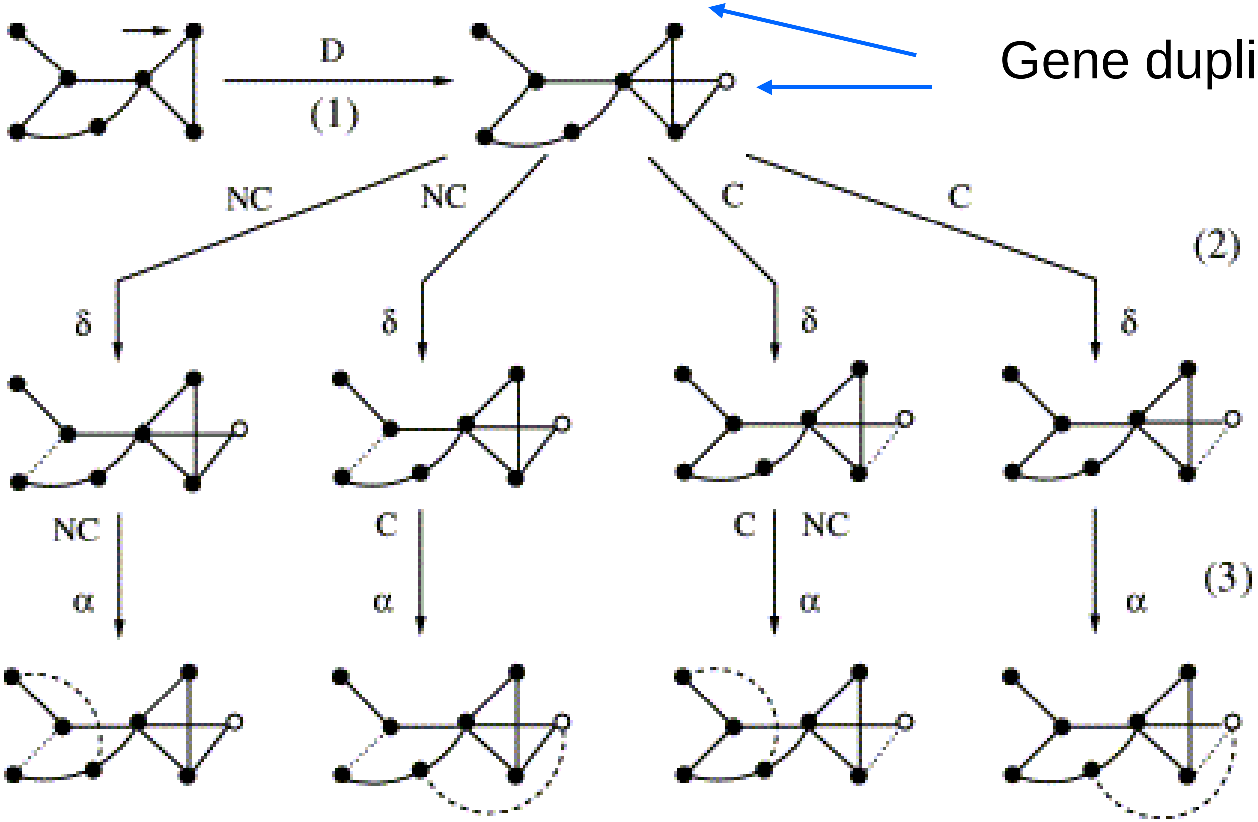
Genes and interactions among gene products have often been conserved through evolution (orthologs).

We can consider the topology of protein interaction networks as a result of a network evolution process.

One can formulate evolving network models for protein interaction networks.

The driving forces behind the formation of edges are gene duplication (i.e. addition of nodes) and mutation (i.e. addition or removal of edges).

Duplication-divergence models



Gene duplication

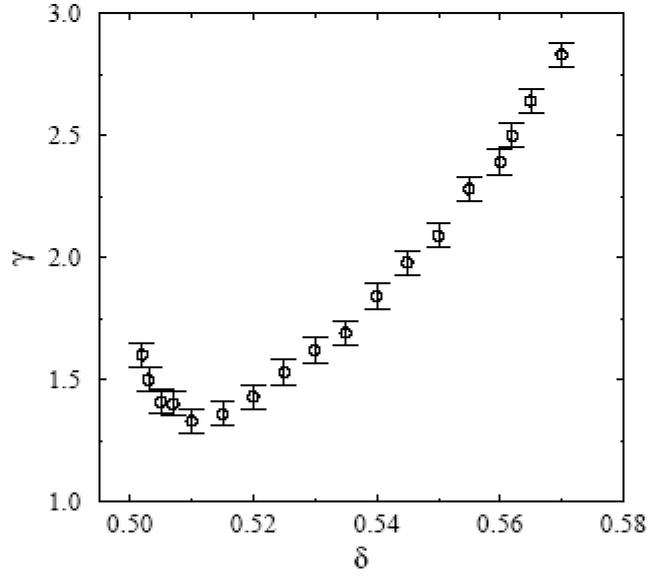
Edge loss with prob. δ

Edge gain with prob. α

Correlated connections (C): only the duplicated gene loses/ gains edge
Uncorrelated connections (NC): edge can be added or removed between any pair of nodes in the network.

Pastor-Satorras et al., Journ. Theor. Biol 22, 199 (2003)

Network properties



Stationary solution possible if $\delta > 0.5$

$$P(k) \propto \left(k_0 + k\right)^{-\gamma} e^{-\frac{k}{k_c}}$$

$$k_c, \gamma = f(\alpha, \delta, N)$$

Apart from a concave region γ is increasing function of δ.

Good agreement with yeast prot. int. network (of N=2000, <K> = 2.5) if δ=0.562

$$\gamma = 2.5 \pm 0.1$$

$k_c \sim 37$

The clustering coefficient of a gene duplication model depends strongly on the initial seed network on which the duplication is performed